BHS Symposium President's Address

How to make advances in hydrological modelling

Keith Beven
Lancaster University

In his invited President's Address, Keith Beven gave a quick summary of his history in hydrological modelling from his first attempt as an undergraduate student in 1970 (modelling the Lynmouth Flood with a programme in Algol that existed on punched cards), through his PhD Finite Element Model of the East Twin catchment in 1975, and his involvement in the development of Topmodel, the SHE model, the IHDM4, Dynamic Topmodel and MIPs models, and for the last 25 years or so, trying to understand uncertainty in hydrological modelling. He noted that we have a real plethora of different hydrological models (from ABC to Xinanjiang, but suggestions are still needed for Y and Z?). Most of those models are conceptual in nature, variants on the stores and storage-flux relationships of the Stanford Watershed Model of the 1960s. Some are based on (incomplete) process theories; some derived directly from the available data (ANN, DBM). Some are distributed and highly integrated into geographical information systems (SHE, Hydrogeosphere, Dynamic Topmodel). Some are modelling systems that allow the user to try out many different structural configurations (Superflex, SUMMA). Some come with databases of parameter values so that they can be applied without calibration (SWAT), although it is usually calibration that allows us to declare some sort of success.

Success after calibration is often illusory, however. Performance often degrades in even the simplest of split-record validation test, and there is no guarantee that a calibrated model will match multiple observables, all parts of the calibration series, or be fit-for-purpose when used to predict the changed conditions often required to provide evidence for decision making. Part of the problem is certainly that the data available for modelling are often not adequate; they can even be hydrologically inconsistent and feed disinformation into the modelling process. But part of the problem is also the lack of rigorous testing and invalidation of models which should best be viewed as multiple working hypotheses about how a catchment or water body functions.

Keith outlined 3 major issues in trying to make advances in hydrological modelling. The first is deciding when a model is not fit-for-purpose and he suggested that what is needed is an appropriate form of Turing Test for models. The second is a better way of representing hydrological processes that, because of the differences in velocities and celerities in the system, has some scale dependence in modelling both flow and transport within a consistent framework. The third is how to manage the "models of everywhere" that are becoming available, with visualisations of model outputs that can be evaluated by local stakeholders.

For the evaluation of models for fitness-for-purpose a Turing Test in the form: "*Can a group of experts tell the difference between a sequence of observations in space and/or time and a model simulation*?" could be considered. Keith noted that we all do this type of evaluation, whether that be in model development and testing, in reviewing papers by other modellers, as clients and consultants in reviewing the results produced by contractors. The question is whether we are sufficiently rigorous in doing so, while recognising the inherent uncertainties in the forcing and evaluation data when some of that data may not be informative in telling us whether a model is fit-for-purpose. A particular point here was the idea of satisfying some statistical criteria 95% (or less) of the time – that still does not necessarily mean that a model will be fit-for-purpose if what we are really interested in (usually flow peaks or low flows) falls within the other 5%. He illustrated the Turing Test concept by an application of the SWAT model to the Morland DTC catchment when it was concluded that SWAT was not fit-for-purpose in that case. He noted, however, that such model failures are actually a good thing – because it means we have to do better.

Fitness-for-purpose is especially important when we wish to predict the impacts of future change. A relevant current example is the need to predict the impacts of Natural flood Management measures to make some assessment of effectiveness when making decisions about investments. Until now, many schemes are being implemented with a belief that there will be adequate benefit, but without such assessments and without any form of monitoring. Predicting the future, however, is subject to important epistemic uncertainties that may be difficult to assess properly, even when that is only how model parameter changes should be implemented.

What an appropriate Turing Test should look like is not yet clear. It will certainly depend on the purpose; it might well depend on what evaluation data are available and assessments of data uncertainties; it might involve decisions about critical experiments or observations that could help in the assessment. Keith used the example of the Blind Validation methodology of Ewen and Parkin (JH, 1996), which was applied to the SHE model in applications to two small catchments as one of the few examples of trying to set criteria for model acceptance before running the model. In both cases the model failed some of the tests, and even the passes were not all that rigorous, but that has not stopped the model being used widely since. This concept is also at the heart of the GLUE Limits of Acceptability methodology that has been used in a number of studies, some of which have rejected even the best models found.

So the next question is whether we can improve our models further. Here Keith suggested that a good way of looking at process representations is closure schemes for the mass and energy balance equations applied at whatever scale of landscape discretisation was appropriate. There is then, however, an implication that process representations should be both hysteretic and scale dependent. Much more needs to be done in understanding what such closure schemes might look like, and he gave the example of using the MIPs particle tracking model, where all the water in the system is tracked as particles that move with different velocities, to investigate changes in hysteresis of the storage-flow relationship at different scales and different antecedent conditions.

Another way in which model improvements might be instigated is by the wider availability of visualisations of model outputs. This provides the means for local stakeholders to relate to what is being predicted in their locality. In the future this will mean much more feedback about where models are wrong, such that modelling will become less the application of models to places and much more a learning process about those places. This raises a variety of research questions, not the least of which is to how to get better data in learning about places (and, again, how to actually do model evaluations both now, and in the future as the system evolves if past simulations can be stored for later access). A final suggestion was that the hydrological community could be much more pro-active in commissioning new measurement techniques.

Keith finished by suggesting that he was finishing his career with much more uncertainty than when he started, but that was a good thing because it meant that there was still so much more to do, especially in reducing epistemic uncertainties particularly through better measurement techniques; in developing better hydrological science as better process representations; and in finding better ways of testing models and being allowed to accept that they might not always fit-for-purpose. A written version of the Address will be included in the special issue of Hydrological Research arising from the BHS Symposium. The slides of the Address may be found at …………….